Secretary of Health and Human Services Advisory Committee on Head Start Accountability and Educational Performance Measures

FINAL REPORT

January 2007

TABLE OF CONTENTS

COMMITTEE MEMBERS

Members
Dr. Don Bailey
Dr. Thomas D. Cook
Dr. Victoria R. Fu
Dr. Vera F. Gutierrez-Clellen
Dr. Ron Haskins
Dr. Susan H. Landry (Chair)
Dr. Christopher J. Lonigan
Dr. Donald A. Rock
Dr. Prentice Starkey
Dr. Dorothy Strickland

Ex-Officio Members
Wade F. Horn, Ph.D.
Jerry Regier

EXECUTIVE SUMMARY

The Advisory Committee on Head Start Accountability and Educational Performance Measures was called by the U.S. Department of Health and Human Services (HHS), Administration for Children and Families (ACF), Office of Head Start. The purpose of the Advisory Committee is to help assess the progress in developing and implementing the Head Start National Reporting System (NRS) and provide recommendations for integrating the NRS with other ongoing assessments of the effectiveness of the Head Start program.

The following key recommendations were made by the Advisory Committee:

RECOMMENDATION 1:
The major purpose of the National Reporting System (NRS) should be to support Head Start programs to provide quality programs that assure children will enter school with age-appropriate skills. To accomplish this goal, clear links need to be specified between technical assistance (TA) approaches and the NRS results. This recommendation includes but is not limited to the development of a clearinghouse of resources, research, and best practices that guide decisions about TA.

RECOMMENDATION 2:
Benchmarks and determination of appropriate gains and outcome levels need to be specified for the NRS to become more effective in meeting intended goals.

RECOMMENDATION 3:
The unit of testing and reporting should be at the child level, and materials and results should be provided to programs in a timely manner. Consideration should be given to allowing the NRS to replace some local assessment approaches.

RECOMMENDATION 4:
Enhancements of the NRS instrument need to continue in order to expand content coverage, provide appropriate training, assure greater reliability and validity, and address issues surrounding children with disabilities.

RECOMMENDATION 5:
Proceed with pilot testing of computer-assisted, adaptive approaches. Piloting of both approaches, computer and Personal Digital Assistant (PDA), should proceed. If both produce equally reliable and valid outcomes, cost should be explored. The use of manipulatives in language and math assessment should be addressed in the pilot test.

The following report is respectfully submitted by the Committee and outlines the above recommendations in detail.

PREAMBLE

Increasingly large numbers of children enter kindergarten without the skills needed to succeed (Zill and West, 2001). This problem is particularly striking for children from low-income households. A growing body of research documents the critical importance of early experiences for learning, and we now understand that high quality early childhood education is an important factor in changing this picture (Bowman, Donovan, & Burns, 2001). For the purpose of this report, high quality is defined as early childhood classrooms that prepare children to enter kindergarten with the full range of skills they need to succeed. This range includes the foundational skills necessary to learn to read, the concepts and skills of elementary school mathematics that permit children to perform well in school, language to express themselves and social competencies. The social competencies necessary for school readiness are well-accepted and include behaviors such as self regulation, social engagement and cooperation. Now there is also mounting research evidence that identifies the early skills that are necessary for success in learning to read: print knowledge, phonological awareness, and language (National Institute for Literacy, 2006). Although less is known about early mathematical development, recent research shows links between the extent of children's early mathematical knowledge and later achievement in mathematics (Starkey & Klein, in press).

A major goal of Head Start is to provide quality programs that assure children will enter school with age-appropriate skills. Thus, there is a clear expectation that children will make good developmental progress with the Head Start intervention. While Head Start children in national studies show cognitive gains from program entry to exit, gains have not been large enough to place the average Head Start child on an equal level with non-poverty kindergarten peers. This may occur in part because some programs are not of sufficient quality to maximize children's growth in language, cognitive, and social skills. Technical assistance and support for local programs are needed to assure that Head Start teachers use effective instructional practices likely to promote school readiness.

In an effort to use research to inform higher quality practice, the Office of Head Start has developed research-based standards that delineate the school readiness outcomes programs need to achieve. These standards address areas such as professional development, instructional practices, and child outcomes. But documenting these standards, while important, is insufficient to ensure full accountability. In the spirit of the Office of Head Start's many new efforts to improve the quality of programs, they also have developed and implemented the National Reporting System (NRS), a valid and reliable child assessment system for assuring that the goal of high quality programs is met. The primary purpose of the NRS is to identify local Head Start programs that need more specialized technical assistance. After this assistance has been provided, there is a need for follow-up to determine whether it was sufficient. Systematic assessment of children's progress and outcomes in Head Start also serves this purpose.

The National Reporting System (NRS) is the first system in the long history of the Head Start program to evaluate the progress, nationwide, of children's skills in the specific developmental areas known to be necessary for school success. Several national databases (Family and Child Experiences Survey, Head Start Impact Study) with extensive information on the performance of

Head Start children were used to develop the NRS, assuring a highly reliable measure. Past and ongoing studies of the NRS demonstrate high levels of reliability and face validity (U.S. Department of Health and Human Services, 2005). Studies of the connection between the NRS battery and children's abilities at the end of kindergarten and teachers' reports of children's academic progress show evidence for the predictive validity of the NRS for measuring the skills and knowledge children need for success in early elementary grades.

The need for the NRS is clear. Without one universal system, it is impossible to reliably identify those programs that need more technical assistance and to tailor such assistance to each program's specific needs. Programs that are not using reliable and valid measures of early literacy, language, math, and social skills are unable to determine if children in their programs have age-appropriate skills or to estimate rates of improvement in skills in relation to their instructional practices. Head Start school readiness behaviors are authentic and legitimate skills, and thus are too important to ignore or only estimate their progress. The NRS has been implemented nationwide to address this need and has the potential to become a unifying force for Head Start programs across the nation to provide quality environments for optimal learning and development.

With any new program there are questions, criticisms, and even fears. Some have questioned the appropriateness of testing 3- and 4-year-olds. Others want more assurance that the NRS is assessing skills in a reliable and valid manner and covers the range of skills and goals that are targeted in the Head Start program. The purpose of the system also has been questioned, with some expressing fears that it will be used to penalize poorly performing programs.

This report and set of recommendations attempt to address many of these concerns. The recommendations emphasize the importance of linking the NRS to effective technical assistance and suggest strategies by which this should be done if programs are going to benefit from the NRS. Careful attention is paid to the appropriate unit of reporting the results to programs so that a balance is achieved between informed planning for professional development and curriculum while also assuring that the NRS is reliable. If specific recommendations covered in this report are addressed, the NRS can produce several benefits. These include measuring a broader range of outcomes, with materials and findings getting to programs in a timely manner, and adequate funds to ensure programs do not compromise their primary mission of assuring children are school ready. The recommendations also address identifying appropriate benchmarks to allow programs to monitor their children's progress. In relation to concerns about having measures that cover a broader range of areas, and thus better reflect the goals of Head Start, recommendations are included addressing this important concern. Expanding and adjusting the developmental domains covered in the NRS should occur in order to improve its utility while always maintaining good reliability and face validity. This might include adjusting or adding items in existing areas of the NRS or including new areas to improve content coverage. Attention also is given to the need to carefully consider how the NRS fits in with other assessments Head Start programs use. The committee discussed the feasibility of the NRS replacing some of the currently used assessment approaches in order to reduce duplication and over-testing children in Head Start classrooms.

We are confident that the NRS can reliably identify settings that are not achieving the full intent of Head Start and inform the implementation of the technical assistance to help programs achieve Head Start's goals. Finally, while some question the appropriateness of assessing young children's learning, evaluation of children's development has been a well accepted practice for decades with developmental pioneers such as Gesell and Bayley and their standardized assessment measures of children's development beginning in infancy. Used appropriately, these data provide the foundation for program improvement so that Head Start can assure parents that all children in Head Start receive the quality programs they need and deserve. In fact, it is only with approaches like the NRS that the nation can assure parents and the public that we know their children are being best served by Head Start.

References

Bowman, B. T., Donovan, S., & Burns, M. S. (Eds.). (2001). *Eager to learn: Educating our preschoolers*. Washington, DC: National Academy Press.

National Institute for Literacy. (2006). Early literacy predictors and effective interventions. Unpublished raw data from the National Early Literacy Panel. Louisville, KY: National Center for Family Literacy.

Starkey, P., & Klein, A. (in press). Sociocultural influences on young children's mathematical knowledge. In O.N. Saracho and B. Spodek (Eds.), *Mathematics, science and technology in early childhood education*. *Contemporary perspectives in early childhood education*. Information Age Publishing.

U.S. Department of Education, National Center for Education Statistics, *Entering Kindergarten: A Portrait of American Children When They Begin School: Findings from The Condition of Education 2000,* Nicholas Zill and Jerry West, NCES 2001–035, Washington, DC: U.S. Government Printing Office, 2001.

U.S. Department of Health and Human Services. (2005). *Report to Congress on the Assessment of children in Head Start Programs*.

RECOMMENDATION 1

The major purpose of the National Reporting System (NRS) is to help Head Start provide quality programs that assure children will enter school with age-appropriate skills. To accomplish this clear links need to be created between technical assistance (TA) goals and the NRS results. In addition, a clearinghouse of resources, research, and best practices that guide decisions about TA should be established.

Recommendation 1A: Educate programs about the NRS and how it can facilitate program improvements.

Making good use of assessment results is critical to an effective and comprehensive early childhood program. Typical uses of assessment include: (1) monitoring learner progress; (2) guiding the decision making required to improve programs; (3) identifying children who might benefit from special services; and (4) reporting to and communicating with others, including parents, policy makers, and administrators. In addition to the ongoing, day-by-day assessments that link closely to the early childhood curriculum, there is a growing trend toward the use of child assessments for program accountability. Regardless of its purpose, the use of assessment data is valuable as a tool for program improvement. Nevertheless, it can also be very challenging for those who may not be prepared to interpret or make use of the results. To facilitate this process we make the following recommendations:

Assure that appropriate and accessible materials accompany the assessment results to assist teachers and administrators in their understanding of the results and their ability to make use of them at both the program and classroom levels.

These materials would include:

(1) An explanation of the assessment instruments and what the "scores" represent;
(2) An explanation of how the results might guide the examination of an existing curriculum and inform the process of curriculum selection or development;
(3) An explanation of how the results might inform decision-making about day-by-day instruction (includes ways to integrate needed areas across the entire curriculum without narrowing instruction to the test);
(4) An explanation of how the results might inform the selection of instructional materials;
(5) An explanation of how this assessment relates to other assessment instruments and procedures in use (e.g., screening instruments, diagnostic tests, informal progress monitoring).

Establish a clearinghouse of resources, research, and best practices.

The knowledge base in early childhood literacy education has grown considerably in recent years. Today's early childhood educators need a strong working knowledge of the general principles of young children's development in all areas of learning. They also need a sound operational knowledge of what those principles look like when instruction is effective in

language, literacy, and math. Teachers should apply that knowledge base in order to observe literacy and math instruction, ask good questions, and make informed decisions about materials, instruction, and assessment. Key components of an early literacy curriculum grounded in evidence-based research include: (1) oral language development, which includes vocabulary and listening; (2) understanding of the alphabetic code, which includes phonological/phonemic awareness and knowledge of the alphabet; and (3) knowledge and understanding about print and its use. Each component has been demonstrated, through research, to be a key part of the essential knowledge base of preschool children who are successful literacy learners in the primary grades and beyond.

For many early childhood educators, much of the concern about assessment is grounded in a lack of understanding of its purpose and usefulness for curriculum and instruction. Teachers and administrators wonder how to link the assessment results in literacy to what is known about other aspects of child development. A clearinghouse of resources, research, and best practices could help educators make these links both philosophically and in practical terms. Topics that might be addressed include: keeping play at the forefront of literacy learning, providing a print rich environment, encouraging culturally and linguistically responsive teaching, attending to prevention and intervention, organizing for differentiated instruction, balancing skills and strategies, scaffolding children's learning, using technology wisely, and integrating literature and literacy across the curriculum.

The clearinghouse should include:

- Research results translated into usable information for administrators and teachers;
- Guidelines for selecting curricular programs and materials that are grounded in research findings;
- Specific examples of materials, programs, and activities that may be obtained or observed in person;
- Resources for on site or combinations of on site - on line professional development;
- Suggestions for various types of professional development opportunities, such as teacher study groups and personalized teacher improvement plans.

Recommendation 1B: Link the NRS results to effective technical assistance support.

Technical Assistance (TA) should be closely linked to NRS results. The new TA system now being conducted by the Office of Head Start, which was designed to provide more targeted support that is individualized to grantee's needs, should allow for this link to occur effectively.

Many of the goals of the new TA system have the potential to facilitate more effective links between it and the NRS including a specific focus on deficient grantees, cluster trainings to address needs of a small group of grantees requiring assistance in the same area, and rigorous standards for TA plans that address child outcomes and self assessments.

The new TA system was developed to include a contractor serving each of the twelve regions with a direct link between the TA liaison and the Head Start Manager. Content experts in the fields of literacy, math, child development, etc. in each region would interact with Federal program specialists, thus providing the mechanism by which decisions can be made about the

type of support (based in part on NRS funding) grantees in each region need. Implementation of TA support should be facilitated by close collaboration between the region experts and locally based TA specialists. The goal of the new TA network system to move from a "one-size-fits-all" to a system that allows for individualized support should allow for a greater focus on new and/or deficient grantees.

While this new TA system should greatly facilitate a tighter link between TA support and NRS results for regions and individual Head Start grantees, there are a number of issues that need to be considered for support to be effectively targeted to the NRS results. Currently grantees and local TA specialists do not have clear guidance or training on how to evaluate the TA supports that are needed in light of different NRS profiles. Secondly, protocols need to be developed linking different NRS deficiencies to TA such as professional development and curriculum programs known to be effective in supporting teachers to advance learning in specific skill areas. Recent large-scale experimental studies evaluating the effectiveness of math curricula and classroom organization for math instruction demonstrate strong gains in 3- and 4-year-olds' math development with highly specified teacher actions implemented in small groups with repetition across each week. Positive outcomes for development of print knowledge and phonological awareness also occurs with highly focused, short, small group activities conducted three to four times each week. Thus, it will be necessary for the TA system to take advantage of what we know works from solidly conducted evaluation studies.

When programs have deficiencies in particular child outcomes as measured by the NRS, a "what works" set of professional development and curriculum approaches need to be implemented with available TA funds. This will require development of a list of effective programs and strategies (see Recommendation on Clearinghouse 1A) and national and regional trainers to train teachers to effectively implement the programs. Local TA specialists also will need to be trained so that they can support and evaluate fidelity of implementation of the programs and strategies. Information from this fidelity evaluation should be documented in regional or national databases and used to inform further refinements of TA/NRS links in relation to grantee characteristics (i.e., teacher education and training background, motivation to improve, teaching beliefs, rural vs. urban, program size, director quality of leadership).

Development of highly sensitive TA will only occur if the many variables known to relate to program success are taken into account. The appointment of an advisory body to provide advice on the development of a "what works" clearinghouse that directly links to the skill domains targeted in the NRS is recommended. It is also recommended that there be a specific process by which regional TA providers are alerted when grantees don't meet benchmarks and gains so that the process described above begins to be implemented in a timely manner. Adequate funding is required to develop the necessary capacity to support programs, including covering all NRS content areas. Currently, expertise in areas such as math and English Language Learners (ELL) is not adequate.

Implementation of Recommendation 1B: The following uses the domain of mathematics to illustrate how NRS findings might be linked to training and technical assistance (TA) if an area of weakness is identified. Depending on the area that needs attention, TA might be targeted more narrowly than what is described below for math. For example, in the literacy domain the

TA might be specific to a weakness in the domain element of phonological awareness and thus might trigger a sequence of steps specifically aimed at strengthening program support for this one literacy area.

Mathematics Example:

Step 1. NRS mathematics outcomes reveal weakness in a local program. The weakness is identified and triggers instructional improvement planning.
Step 2. Instructional improvement planning begins in the Office of Head Start (OHS) by accessing OHS records to determine whether the NRS mathematics outcome is a recurring or new weakness in the program.

(a) If this is a recurring weakness, records are accessed to determine what measures were taken previously to strengthen the program.

(b) If the weakness is new, records are accessed to determine what was found during the most recent (PRISM)[1] review of this program pertaining to mathematics.

(c) OHS records are also accessed to provide pertinent background data, such as the identity of the curriculum used by the program to support children's mathematical development, the quality of curriculum implementation (if known), the amount of professional development in mathematics that teachers or supervisors in the program have received and when they last received it, and other data deemed pertinent.

Step 3. The Office of Head Start notifies the grantee and the appropriate regional office that the NRS has identified a weakness in math and that instructional improvement planning should occur. Information gathered by the Office of Head Start in Step 2 (above) is provided to the grantee and regional office to inform the planning process.
Step 4. A mathematics content specialist in the regional office initiates contact with the grantee to discuss the finding and relevant information gathered in Step 2. The scope of the weakness will be ascertained by examining NRS mathematics data at multiple levels. The levels could include classrooms and centers, subgroups of children (e.g., LEP children), the delegate agency or the entire program
Step 5. After the likely moderators of child math outcomes have been identified, types of TA resources (e.g., a new mathematics curriculum, training of some or all supervisors and/or teachers to improve implementation of the current mathematics curriculum) that the grantee needs specified in a Mathematics Instructional Improvement Plan jointly produced by the grantee and regional office.
Step 6. The grantee, in consultation with the regional office, will select specific resources (e.g., a particular math curriculum, a specific type of math training by a qualified trainer) from the Head Start clearinghouse that are needed to implement the Mathematics Instructional Improvement Plan.
Step 7. The Office of Head Start will then approve and earmark TA funds to implement this plan.
Step 8. The instructional improvement plan will be implemented.
Step 9. NRS math outcomes for the next period will be examined for improvement.

For effective mathematics TA to be provided to local programs, three systemic modifications are needed in the TA currently being provided. First, there is a need for systematic alignment of

[1] PRISM: Program Review Instrument for Systems Monitoring

mathematics content throughout the Head Start system. Components needing alignment include (a) the Head Start Child Outcomes Framework's domain elements of mathematics for Head Start, (b) the NRS (across multiple forms), which should assess multiple specific mathematics indicators within each domain element, (c) local mathematics assessments, which should include items that correspond to the specific mathematics indicators, (d) mathematics curricula used by programs, which should include broad support for the acquisition of mathematical knowledge within each element of the domain, and (e) the mathematics section of PRISM (p. C-47, PRISM 2005).

In addition to alignment, PRISM's mathematics examples need a degree of specificity similar to that provided for language and literacy (p. C-45). The Department should consider adding examples such as "learning activities that promote knowledge of mathematics within each domain element." Alignment of mathematics in Head Start and Early Head Start should be considered if the Office of Head Start develops the capacity of Early Head Start programs to support the development of early mathematical knowledge as a distinct area of cognitive development. Domain elements of mathematics of Head Start and Early Head Start should be aligned when possible.

Second, development of the capacity of Head Start to provide programs with quality TA support in mathematics is needed. Specifically, (a) regional offices should have content specialists in mathematics, (b) regional TA contractors should be certified for expertise in providing mathematics TA, and (c) local education coordinators, supervisors, and teachers should be provided with training in children's mathematical development and how learning environments can support it.

Third, a sound procedure will be needed for identification of resources to strengthen local programs in the content area of mathematics. These resources might include (a) aligned, evidence-based mathematics curricula, (b) instruments to measure the fidelity of curriculum implementation, (c) aligned instruments for supplementary local assessments of children's mathematical knowledge, (d) aligned instruments for supplementary local assessments of classroom mathematics practices, and (e) qualified mathematics TA providers.

RECOMMENDATION 2

Benchmarks and determination of appropriate yearly test score gains and outcome levels need to be specified for the NRS to become more effective in meeting intended goals.

The identification of Head Start programs and their classrooms that may need additional technical assistance centers on the use of the NRS to (1) identify programs and classrooms that need additional help, and (2) provide limited proscriptive information at the classroom level with respect to which content strands are in greatest need of improvement.

Identification of poorly performing programs

Recommendation: Two criteria should be used for identifying poorly performing programs: (1) performance at or above a criterion-referenced cutting score, and (2) failure to achieve an appropriate rate of gain.

The use of only one of these criteria may not adequately demonstrate how well children are doing in a particular program. A program would be identified for TA if it did not meet either of the two criteria. The first criterion is somewhat more empirically driven than is the second criterion. Using the FACES database one could use the NRS cognitive scores in the four strands (language, literacy, math and social) to develop an equation for predicting whether a child would be promoted to second grade. As we understand it, the FACES database has both NRS scores on completion of Head Start and information on whether the same children successfully completed first grade. Since the outcome here is categorical (promoted or not) a logistic regression equation would be the most appropriate statistical procedure. Then using the prediction equation based on the FACES data, the four cognitive score means from each **classroom** would be entered into the equation and a predicted probability of say .66 or greater would be considered acceptable. That is, the average child in a given classroom having completed the Head Start program should have 2 out of 3 chances of successfully completing first grade. At this point the selection of a probability of .66 is relatively arbitrary and may be modified when the distributions of pass-fail and the empirical predicted probabilities are examined more closely. This prediction model assumes, of course, that the NRS cognitive scores are significant predictors of whether the child passes first grade.

The second criterion is based on the average gains associated with each classroom. This second "gain" criterion is necessary because it is possible that in some classrooms the cognitive level of the children at entry may be so low that even with quite large gains they might not attain sufficient increases in their cognitive performance to satisfy the first criterion. That is, even with relatively large gains in a given classroom suggesting that the instructional program is doing well, the average child's performance in the particular classroom might not attain the levels on the cognitive measures necessary to yield a predictive probability of .66. The question now is what is a satisfactory criterion of gain? It is suggested here that a good starting point might be a mean classroom gain in standard deviation units equivalent to that found for the total Head Start population.

The gain criterion could also be formulated in terms of percentiles. That is, a classroom's mean gain must be at or above the 50th percentile. This classroom mean gain criterion would be completed separately for each of the four content strands within each classroom. Both the logistic regression equation and the gain score norms could be based on the previous year's population. This approach would allow for quick feedback to program administrators in the sense that the reporting would not be held up waiting on the current-year scores. As in the case of the first criterion, further empirical analysis of distributions of gains for selected subpopulations from available databases could provide further evidence to support or deny the validity and practicality of these initial cutting-score criteria.

It should be kept in mind that programs and classrooms to be selected for additional help would have to under perform on both criteria. That is, if the analysis of the data from a particular classroom suggests that the average child at exit has less than a 2 out of 3 chance of passing first grade **and** the average child in that classroom also did not meet the gain criterion on any **one** of the four content strands, then that classroom would be identified for further mentoring. The TA would only be given in the content area that did not meet the gain criteria.

The gain criterion assumes a pre- and post-testing and that the content strand performance is either based on an adaptive testing approach (discussed under Recommendation 5) or steps have been taken to ensure no floor and ceiling effects. While the cutting scores suggested here seem reasonable, a cost utility model based on a range of cutting scores might be posed. In the real world there is only so much money available for additional mentoring and the setting of cutting scores and growth rates that are excessively high might identify far more classrooms in need of mentoring than the budget would allow.

The above discussion has not mentioned any special needs sub-populations and their relationship to the definition of the two performance criteria outlined above. It is our position that all programs should be held to the same criteria. However, one of the reasons that two criteria are being proposed is that some special needs sub-populations (e.g., children with Individualized Education Plans and English Language Learner) could perform quite poorly at entry but we would expect them to at least show growth rates at or above the 50th percentile based on the previous year's data.

RECOMMENDATION 3

The unit of testing and reporting should be at the child level and materials and results should be provided to programs in a timely manner. Consideration should be given to allowing the NRS to replace some local assessment approaches.

Recommendation 3A: At least temporarily the NRS should continue to be administered to all students rather than to a sample of students from each program.

Although the major purpose of the NRS is to identify programs that need technical assistance in order to boost student performance, the NRS can be useful at the program, center, classroom, and individual level. The major reason we are recommending that the NRS continue to be administered to all students is that the test cannot be useful in assessing the progress of individual students nor of classrooms unless all students receive the test.

As the NRS is administered year after year and the results published, the average performance of programs at the national, regional, state, and sub-state levels will become available. Thus, it will be possible for administrators at the program and center levels to compare their performance with that of other programs across the country and across their state and region. This information will be valuable to administrators who want to focus their attention and resources on centers and classrooms that seem to be underperforming. Similarly, administrators can use the results to trace changes over time in order to take appropriate action.

NRS results can also be used as a rough measure of the performance of individual students. The test yields scores in language, vocabulary, pre-reading, and math that allow comparisons with performance of other preschoolers and with national norms. The reliability level of the test, roughly in the range of .8, means it typically yields an accurate measure of student learning in these four areas. The Committee is aware that the NRS results should not be used as the sole measure of student learning. Rather, it must be used in combination with other assessments, especially the direct observation of student learning and behavior by the classroom teacher. Even so, when used as one element of a comprehensive assessment approach, the NRS can make a valuable contribution to understanding the performance of individual students and ensuring that they are receiving the experiences and instruction they need to be adequately prepared for kindergarten. The test can also signal that an individual student may have serious learning problems, in which case further assessment and diagnosis may be appropriate.

Recommendation 3B: Local programs should consider whether the NRS can be used in place of locally developed tests for purposes of planning instruction and identifying students with unexpectedly low scores.

The reliability of the test is sufficient to yield scores that are likely to be stable and accurate, a feature that many local tests either do not posses or is unknown. In addition, it is advantageous for administrators and teachers to know how their students are performing relative to similar students, a feature of the NRS that is quite unlikely to characterize local tests. Each of the four

NRS subtests yield estimates of five levels of student proficiency, thereby helping the classroom teacher or local administrator gauge the level of student performance relative to that of other students.

Recommendation 3C: The Department and its contractor should take two actions to ensure that the tests can be effectively used by local programs to improve instruction: timely receipt of results and instructions on how programs can use the test results to improve actions.

For maximum usefulness in improving instruction, test results must be reported to individual programs in a timely fashion. The contractor conducting NRS testing has assured the Advisory Council that it is working to provide test results to programs sooner. Wider use of computer-assisted testing would allow the contractor to provide results that allow comparison of performance by classrooms and individual students almost immediately. The Department should take steps to ensure that program administrators and teachers at the local level are offered instruction in how to use NRS results to improve teaching and outcomes.

RECOMMENDATION 4

Enhancements of the NRS instrument need to continue in order to expand content coverage, provide appropriate training, assure greater reliability and validity, and address issues surrounding children with disabilities.

Recommendation 4A: If a child fails the English testing, the center should use local resources to assess the child in the home language.

Children from diverse backgrounds must be tested using culturally and linguistically appropriate means. Failure to do so underestimates their true skills and progress in the program. Currently, the NRS focuses only on English or Spanish speakers, and a large number of English language learners are not evaluated in their home languages. For these children, no information regarding their readiness or needs is available, based on the current state of development of the NRS. This gap in knowledge is critical for the identification of English learners with language disabilities who need early intervention services. To appropriately identify these children, their skills need to be assessed in both languages.

Recommendation 4B: Provide alternative Spanish receptive vocabulary items to address dialectal differences across the typical Spanish varieties spoken in the U.S.

One of the domains with the potential to show growth in the languages of ELLs is vocabulary. However, the current receptive vocabulary measure of the NRS does not adequately address differences across children groups. The English vocabulary items do not appear to penalize dialectal or regional differences, yet concerns have been raised about the Spanish vocabulary measure. It is suggested that the NRS Spanish Vocabulary be refined to include alternative responses that are appropriate across Puerto Rican and Mexican American speakers.

Recommendation 4C: Study the validity of using a composite score that credits the child's use of the two languages in their responses to the test.

Children who are bilingual may know certain items in one language and not in the other. For example, the child may know certain vocabulary words in Spanish and other words in English; numbers and letters in English and not in Spanish; or math concepts in Spanish but not in English. Current research in the area of assessment of bilingual preschoolers indicates that vocabulary testing should consider these responses as part of one conceptual system. A composite score should reflect the child's true competencies. The scoring of the vocabulary section should not penalize code switching.

Recommendation 4D: Administer the two languages of the NRS on separate testing dates or at different times.

Training of bilingual assessors is critical for obtaining valid and reliable data from English Language Learners. This training should include administration of the NRS in separate testing dates (one for each language); certified language proficiency of the assessors; and a system for

ensuring monitoring of test administration. When English Language Learners are tested in both languages, test performance can be affected by the child's level of fatigue.

Recommendation 4E: Develop and study the validity of a measure of phonemic awareness to be included in the NRS.

Research has demonstrated that the domain of phonological awareness can provide important information regarding a child's reading readiness. Initial work evaluating phonological awareness within the NRS has not resulted in adequate levels of inter-tester reliability. Further development and evaluation is needed to determine the predictive value of this measure during the preschool years with both monolingual and English Language Learners.

Recommendation 4F: Construct the mathematics subtest in a manner that ensures good content validity.

Specifically, NRS items in the domain of mathematics should correspond to *domain elements* and their *indicators* included in the current version of the Head Start Child Outcomes Framework. Particular *indicators* of children's mathematical skills and knowledge, as represented by NRS items, should be varied across forms of the NRS. The rationale is to determine whether children are acquiring a broad foundation of mathematical knowledge.

The current version of the NRS is especially in need of further development to assess knowledge of patterns and to broaden its assessment of spatial knowledge.

Utilization of manipulatives for some tasks should also be considered to determine whether manipulation provides a more sensitive measure of children's knowledge.

Recommendation 4G: Head Start has a mandate to serve children with disabilities and their performance in Head Start needs to be adequately documented. The Committee has developed the following set of recommendations around the NRS that support this mandate:

1. Children with disabilities should not be excluded from the NRS, unless specifically prohibited by the child's Individual Education Plan;
2. Appropriate accommodation should be developed and tested to maximize the participation of children with disabilities in the NRS;
3. Teachers should be given adequate training and guidance in assessing children with disabilities;
4. Children with disabilities should be included in the social-emotional assessments;
5. Performance of children with disabilities should be reported separately from the performance of children without disabilities
6. The Office of Head Start should develop a plan for evaluating, at the local and national level, the extent to which Head Start is adequately supporting the development of children with disabilities.
7. Efforts to measure progress and outcomes for children with disabilities in Head Start should be aligned with efforts to measure progress and outcomes for children with

disabilities of comparable age served by the U.S. Department of Education under the Individuals with Disabilities Education Act.

Since 1972, Head Start has been required to have at least 10% of its enrollees be children with disabilities. Head Start has consistently met or exceeded this goal, with an average of approximately 14% of recent enrollment comprising children with disabilities. The most common disability is speech and hearing impairment, but other disabilities are also represented.

The NRS has included children with disabilities in its assessments, but several challenges have emerged. These include:

- Under-representation of children with disabilities in the reporting sample (e.g., in the Fall 2003 assessment only 7% of the sample had a disability).
- Uncertainty remains among programs with regard to whether all children with disabilities should be assessed, whether accommodations can be made, and if so, which accommodations are allowable.
- Many teachers do not feel adequately trained to assess children with disabilities.
- Data suggest that children with disabilities are further behind in vocabulary, math, and letter-naming skills at the end of program year than children without disabilities.
- Little work has been done to explore the usefulness or applicability of newly proposed measures (e.g., socio-emotional measures) for children with disabilities.
- The nature and impact of disabilities vary widely across children, making it difficult to develop adaptations appropriate to every child and to determine whether growth trajectories are adequate.
- Any given classroom will only serve a small number of children with disabilities, thus making it difficult to conclude with confidence whether, at the classroom level, the program is being effective for children with disabilities.

Recommendation 4H: Eliminate the health measures from the NRS.

Attempts have been made to enhance the NRS instrument by adding health measures (e.g., weight, vaccinations, etc). However, these measures duplicate existing activities that programs already have in place. Programs engage in health screenings as part of their local assessments.

RECOMMENDATION 5

Proceed with pilot testing of computer-assisted, adaptive approaches. Piloting of both approaches, computer and Personal Digital Assistant, should proceed. If both produce equally reliable and valid outcomes, cost should be explored. The use of manipulatives in language and math assessment should be addressed in the pilot test.

Adaptive Tests, Behavioral Anchoring, and Measuring Change with the NRS

Adaptive tests with behavioral anchoring are necessary for measuring change in order to estimate the "value added" contribution of attending Head Start as well as providing diagnostic information at the classroom level. Adaptive testing procedures provide a means for selecting items from the total item pool in such a way as to match the item difficulties to each individual child's ability level. This tailoring of the items to any given child's specific ability level minimizes potential frustration of the child with respect to the testing task and also provides more accurate estimates within a shorter testing time. While different children will be receiving at least some items that may be different from their peers, the items can be put on the same scale using Item Response Theory (Lord, 1980; Embretson & Reise, 2000).

The use of adaptive tests for measuring gains is even more critical because:

(1) Adaptive tests will lead to more accurate estimates as well as unbiased estimates of gains, i.e.:

a. Standard errors of measurement are stable and consistently high across all values on the vertical score scale. Gains at the low or the high end of the score distribution are measured with almost the same precision as gains in the middle of the test score distribution.

b. Floor and ceiling effects will be minimized, which is absolutely essential for measuring change and even more so in a value-added context in order to avoid biased estimates. Without the broad range ability assessment capability of adaptive tests, gains at both the low and high end of the test score scale tend to be underestimated (biased).

(2) Behavioral anchoring the vertical test score scale is desirable for measuring change because:

a. Children entering Head Start at the national level will be characterized by a relatively wide range of ability and preparation and thus will be changing at different places along the vertical scale. In addition to knowing how much change is taking place we also need to know at what skill level the growth is taking place. Children in classrooms who on average are changing at the upper end of the scale are learning somewhat different skills than those children making equivalent amounts of change at the lower end of the scale. The behavioral anchors provide descriptions of the skills involved at various score points along the vertical test scale.

b. The use of the behavioral anchors in conjunction with adaptive testing procedures would provide classroom teachers and program administrators with accurate estimates of the amount of gain as well as some diagnostic information about what kinds of growth are occurring.

Recommendation 5A: The NRS is now using adaptive methods in vocabulary and letters but not in the remaining two cognitive strands; adaptive testing should be expanded to the remaining two strands.

Recommendation 5B: The NRS has a limited number of behavioral anchors; they should be reviewed and possibly be expanded.

Recommendation 5C: In addition to the amount of gain reported at the classroom level, the NRS should report the skill level at which the typical child in a given classroom is making gains.

References

Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Lord, F. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Other Issues the Committee Considered

A. Need for Additional Resources

Recommendation: Consideration might be given to areas that may be in need of additional resources.

Candidates for supplementary funding might include activities associated with the administration of the NRS, increased technical assistance to help teachers and administrators interpret and make effective use of test results, and special initiatives designed to help parents understand the purpose and use of test results.

It is hoped that the NRS would eventually become the centerpiece of a well coordinated program of assessment and professional development with appropriate links to other existing sources of data, such as health and safety information and parent programs.

B. Families of Children in Head Start

From its inception, Head Start has had a primary goal of helping children from low-income families succeed in school. The NRS has focused exclusively on assessments related to child outcomes of Head Start. But a wealth of literature suggests that families also can benefit from programs such as Head Start, and indeed supporting families has always been part of the Head Start mission and philosophy. In 1995, Head Start began a process of identifying a set of program performance measures, with five broad objectives and a set of accompanying indicators. One of the five objectives directly addresses family outcomes—to strengthen families as the primary nurturers of their children, with three indicators:

- Head Start parents demonstrate improved parenting skills.
- Head Start parents improve their self-concept and emotional well-being.
- Head Start parents make progress toward their educational, literacy, and employment goals.

A second goal indirectly addresses family outcomes—to link children and families to needed community services, with four indicators:

- Head Start parents link with social service agencies to obtain needed services.
- Head Start parents link with education agencies to obtain needed services.
- Head Start parents link with health care services to obtain needed care.
- Head Start parents secure child care in order to work, go to school, or gain employment training.

Although some data was collected on family outcomes as part of an interview conducted in the Family and Child Experiences Survey study, family outcomes have not been the focus of the NRS.

Recommendation: The Committee feels that family outcomes of Head Start need to be documented, but that this work should not be a part of the NRS at the present time; rather the Office of Head Start might consider beginning a parallel set of activities to identify or develop tools for documenting family outcomes and begin pilot studies with these measures.

Three ultimate goals should be met by this work: (1) a national picture of the extent to which Head Start benefits families, (2) data on which programs might need technical assistance to maximize family benefit; and (3) a technical assistance component that identifies best practices in working with families and provides support to local programs in order to maximize family benefit.